Vol. 6. No. 2. November 2024
P-ISSN : 2714-5789 | E-ISSN : 2986-7959
Jurnal Sains dan Sistem Teknologi Informasi (SANDI)

DATA MINING KLASTERISASI CUSTOMER SEGMANTATION NETFLIX MENGGUNAKAN METODE K-MEANS DENGAN RAPIDMINER

[1)]Nur Maymuna, [2)]Zaehol Fatah
[1]Sistem Informasi, Universitas Ibrahimy
[2]Sistem Informasi, Universitas Ibrahimy
[1]nurmaymuna52@gmail.com , [2]zaeholfatah@gmail.com

ABSTRACT

The intense competition in the streaming industry has pushed Netflix to understand its customer base more deeply. This study aims to segment Netflix customers using the K-Means Clustering method to identify customer characteristics and behavior patterns. The analysis was carried out using the RapidMiner Studio tool, utilizing the Netflix Userbase dataset for the period June-July 2023, which consisted of 2500 customer data. The variables used include Subscription Type, Monthly Revenue, Country, Age, Gender, and Device. The results of implementing K-Means Clustering with k = 5 produced optimal customer segmentation, as evidenced by the Davies Bouldin Index value of -2.920. The analysis identified five main clusters: new customers with high engagement levels (Cluster 0), loyal old customers (Cluster 1), active customers with regular payment patterns (Cluster 2), dominant customers from the United States (Cluster 3), and a special segment of German customers (Cluster 4). The model evaluation shows good cluster cohesion with optimal average within centroid distance values for each cluster. This study recommends Netflix to develop a more personalized marketing strategy based on the characteristics of each cluster, and to conduct segmentation analysis periodically to follow changes in customer preferences. The results of this segmentation can be the basis for strategic decision-making in developing more targeted services and marketing.

Keywords: *Netflix, Segmentation Users, K-Means Clustering, RapidMiner*

ABSTRAK

Persaingan ketat dalam industri streaming mendorong Netflix untuk memahami basis pelanggannya secara lebih mendalam. Penelitian ini bertujuan untuk melakukan segmentasi pelanggan Netflix menggunakan metode K-Means Clustering untuk mengidentifikasi karakteristik dan pola perilaku pelanggan. Dengan memanfaatkan dataset Netflix Userbase periode Juni-Juli 2023 yang terdiri dari 2500 data pelanggan, analisis dilakukan menggunakan tools RapidMiner Studio. Variabel yang digunakan meliputi Subscription Type, Monthly Revenue, Country, Age, Gender, dan Device. Hasil implementasi K-Means Clustering dengan k=5 menghasilkan segmentasi pelanggan yang optimal, dibuktikan dengan nilai Davies Bouldin Index -2.920. Analisis mengidentifikasi lima cluster utama: pelanggan baru dengan tingkat engagement tinggi (Cluster 0), pelanggan lama yang loyal (Cluster 1), pelanggan aktif dengan pola pembayaran teratur (Cluster 2), pelanggan dominan dari United States (Cluster 3), dan segmen khusus pelanggan Germany (Cluster 4). Evaluasi model menunjukkan kohesi cluster yang baik dengan nilai average within centroid distance yang optimal untuk setiap cluster. Penelitian ini merekomendasikan Netflix untuk mengembangkan strategi pemasaran yang lebih personal berdasarkan karakteristik masing-masing cluster, serta melakukan analisis segmentasi secara berkala untuk mengikuti perubahan preferensi pelanggan. Hasil segmentasi ini dapat menjadi dasar pengambilan keputusan strategis dalam pengembangan layanan dan pemasaran yang lebih terarah. cluster, serta melakukan analisis segmentasi secara berkala untuk mengikuti perubahan preferensi pelanggan.

Kata Kunci: *Netflix, Segmentasi Pelanggan, K-Means Clustering, RapidMiner*

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024

I. PENDAHULUAN

Perkembangan teknologi digital yang pesat telah mengubah cara masyarakat mengkonsumsi konten hiburan. Salah satu bentuk transformasi tersebut adalah berkembangnya layanan streaming video yang memungkinkan pengguna menonton konten kapanpun dan dimanapun. Netflix, sebagai salah satu pionir layanan streaming, telah menjadi pemain dominan dalam industri ini sejak didirikan oleh Reed Hastings dan Marc Randolph pada tahun 1997 [1]. Platform ini telah bertransformasi dari layanan penyewaan DVD melalui pos menjadi layanan streaming global dengan lebih dari 247 juta pelanggan di seluruh dunia per tahun 2023. Seiring dengan pertumbuhan industri streaming, persaingan menjadi semakin ketat dengan hadirnya berbagai kompetitor seperti Disney+, Amazon Prime Video, dan HBO Max. Situasi ini mendorong Netflix untuk terus berinovasi dan memahami karakteristik pelanggannya dengan lebih baik. Salah satu pendekatan yang dapat digunakan untuk memahami basis pelanggan adalah melalui teknik data mining, khususnya analisis klasterisasi pelanggan (customer segmentation) [2].

Data mining merupakan proses ekstraksi pola dan pengetahuan dari sejumlah besar data yang dapat memberikan wawasan berharga bagi pengambilan keputusan bisnis. Dalam konteks Netflix, teknik clustering khususnya algoritma K-Means dapat digunakan untuk mengidentifikasi kelompok-kelompok pelanggan dengan karakteristik serupa berdasarkan berbagai variabel seperti jenis langganan, pendapatan bulanan, demografis, dan perilaku menonton. Metode K-Means dipilih karena kemampuannya dalam memproses data numerik dengan efisien dan menghasilkan cluster yang mudah diinterpretasi [3].

RapidMiner sebagai salah satu tools data mining yang populer menyediakan antarmuka visual yang memudahkan proses analisis klasterisasi. Platform ini mendukung berbagai algoritma machine learning termasuk K-Means dan menyediakan fitur visualisasi yang membantu dalam interpretasi hasil [4]. Dengan menggunakan dataset Netflix Userbase yang tersedia di Kaggle yang mencakup 2500 data pelanggan dengan berbagai atribut seperti tipe langganan, pendapatan bulanan, negara, usia, gender, dan perangkat yang digunakan, penelitian ini bertujuan untuk mengidentifikasi segmen-segmen pelanggan Netflix yang dapat digunakan sebagai dasar pengembangan strategi bisnis yang lebih efektif.

Hasil segmentasi pelanggan ini diharapkan dapat membantu Netflix dalam mengoptimalkan strategi pemasaran, pengembangan konten, dan peningkatan layanan yang sesuai dengan karakteristik masing-masing segmen pelanggan. Selain itu, pemahaman yang lebih baik tentang basis pelanggan dapat mendukung Netflix dalam mempertahankan posisinya sebagai pemimpin pasar di industri streaming yang semakin kompetitif [5].

Penelitian ini bertujuan untuk: 1. Mengimplementasikan algoritma K-Means dalam proses segmentasi pelanggan Netflix, 2. Menganalisis karakteristik setiap segmen pelanggan yang terbentuk, 3. Memberikan rekomendasi strategi bisnis berdasarkan hasil segmentasi. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi praktis bagi pengembangan strategi bisnis Netflix serta kontribusi teoretis dalam pengembangan metode segmentasi pelanggan menggunakan teknik data mining.

II. METODE PENELITIAN

2.1 Netflix

Netflix merupakan platform streaming yang menyediakan layanan konten digital berbasis langganan (subscription-based). Didirikan pada tahun 1997, Netflix telah berevolusi dari layanan penyewaan DVD menjadi platform streaming global terkemuka. Platform ini menawarkan tiga jenis paket berlangganan yaitu Basic, Standard, dan Premium, dengan fitur dan harga yang berbeda-beda. Diferensiasi paket ini memungkinkan Netflix untuk melayani berbagai segmen pelanggan dengan preferensi dan kemampuan finansial yang beragam. Model bisnis Netflix berfokus pada penyediaan konten original dan lisensi, didukung oleh sistem rekomendasi berbasis algoritma untuk meningkatkan pengalaman pengguna [6].

2.2 K-Means Clustering

K-Means Clustering adalah salah satu metode unsupervised learning yang bertujuan untuk mempartisi data ke dalam K kelompok (cluster) berdasarkan karakteristik yang sama. Algoritma ini bekerja dengan menentukan K titik pusat (centroid) secara acak, kemudian mengelompokkan setiap data ke centroid terdekat menggunakan perhitungan jarak Euclidean. Proses ini diulang dengan memperbarui posisi centroid berdasarkan rata-rata anggota cluster hingga mencapai konvergensi. Kelebihan K-Means terletak pada kesederhanaan implementasi dan efisiensi komputasi, terutama untuk dataset berskala besar. Namun, penentuan jumlah cluster optimal (K) merupakan tantangan utama yang biasanya diatasi dengan metode elbow atau silhouette analysis [7].

2.3 Segmentasi Pengguna

Segmentasi pengguna adalah proses membagi basis pelanggan menjadi kelompok-kelompok dengan karakteristik, kebutuhan, atau perilaku yang serupa.

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024

Dalam konteks layanan streaming seperti Netflix, segmentasi dapat dilakukan berdasarkan berbagai variabel seperti demografis (usia, gender, lokasi), behavior (frekuensi menonton, genre favorit), dan ekonomi (jenis langganan, pendapatan) [8].

2.4 Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan klasterisasi segmentasi pelanggan Netflix. Proses penelitian dibagi menjadi empat tahapan utama [9]:

a. Pengumpulan Data

Penelitian ini menggunakan dataset Netflix Userbase dari platform Kaggle yang terdiri dari 2500 sampel data pelanggan dengan 10 variabel meliputi User ID, Subscription Type, Monthly Revenue, Join Date, Last Payment Date, Country, Age, Gender, Device, dan Plan Duration.

b. Preprocessing Data

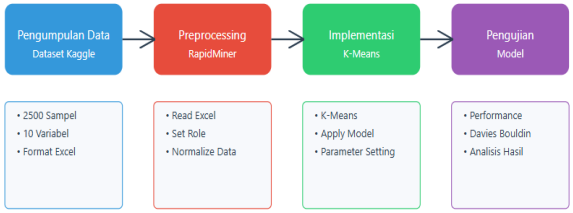
Tahap preprocessing dilakukan menggunakan tools RapidMiner Studio, meliputi beberapa langkah: 1. Read Excel untuk mengimpor dataset. 2. Set Role untuk menentukan atribut yang akan digunakan dalam proses clustering. 3. Normalisasi data untuk menyeragamkan skala nilai pada setiap atribut. Proses ini penting untuk memastikan kualitas data yang akan diolah.

c. Implementasi K-Means

Pada tahap ini, algoritma K-Means diterapkan dengan menentukan jumlah cluster optimal menggunakan metode elbow. Parameter yang digunakan meliputi measure types euclidean distance dan max runs 10 untuk memastikan konvergensi hasil clustering. Proses ini menghasilkan pengelompokan pelanggan berdasarkan karakteristik yang serupa.

d. Pengujian Model

Evaluasi hasil clustering dilakukan dengan menggunakan metrik Davies Bouldin Index untuk mengukur validitas cluster yang terbentuk. Selanjutnya dilakukan analisis karakteristik setiap cluster untuk memberikan interpretasi dan rekomendasi bisnis yang relevan bagi Netflix [10].



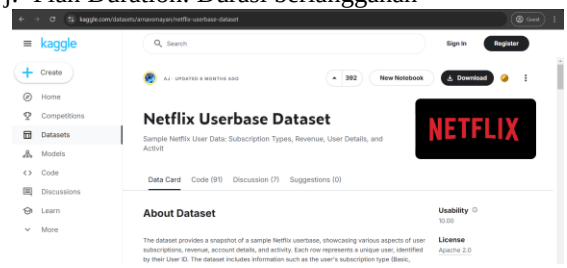
GAMBAR 1.
METODOLOGI PENELITIAN

III. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini adalah Netflix Userbase yang diambil dari platform Kaggle, mencakup periode Juni-Juli 2023 dengan total 2500 data pelanggan Netflix. Dataset ini terdiri dari 10 variabel yang merepresentasikan karakteristik pelanggan Netflix, yaitu:

a. User ID: Nomor identifikasi unik pelanggan
b. Subscription Type: Jenis langganan (Basic, Standard, Premium)
c. Monthly Revenue: Pendapatan bulanan dalam dollar ($10-15)
d. Join Date: Tanggal bergabung pelanggan
e. Last Payment Date: Tanggal pembayaran terakhir
f. Country: Negara asal pelanggan
g. Age: Usia pelanggan (rentang 26-51 tahun)
h. Gender: Jenis kelamin (Male, Female)
i. Device: Perangkat yang digunakan (Smartphone, Laptop, Smart TV, Tablet)
j. Plan Duration: Durasi berlangganan



GAMBAR 2.
HALAMAN WEBSITE KAGGLE

Dalam analisis ini, fokus diberikan pada variabel yang relevan untuk segmentasi pelanggan yaitu Subscription Type, Monthly Revenue, Country, Age, Gender, dan Device. Pemilihan variabel ini didasarkan pada pertimbangan bahwa variabel-variabel tersebut dapat memberikan gambaran yang jelas tentang karakteristik dan perilaku pelanggan Netflix. Dataset ini menyediakan informasi yang cukup komprehensif untuk mengidentifikasi pola-pola segmentasi pelanggan yang dapat digunakan dalam pengembangan strategi bisnis yang lebih terarah.

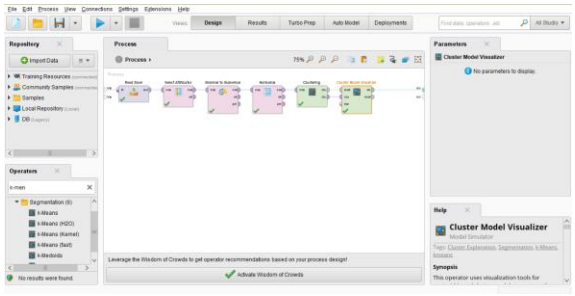
3.2 Preprocessing Data

Tahapan preprocessing data dalam penelitian ini dilakukan menggunakan RapidMiner Studio dengan beberapa langkah yang saling berkaitan. Pertama, data Netflix Userbase diimpor menggunakan operator "Read Excel". Kemudian dilakukan seleksi atribut menggunakan operator "Select Attributes" untuk memilih variabel yang relevan dalam analisis clustering, yaitu Subscription Type, Monthly

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024

Revenue, Country, Age, Gender, dan Device. Atribut seperti Join Date dan Last Payment Date tidak diikutsertakan karena kurang relevan untuk segmentasi pelanggan.

Selanjutnya, dilakukan transformasi data kategorikal menjadi numerik menggunakan operator "Nominal to Numerical" untuk atribut Subscription Type, Country, Gender, dan Device. Hal ini diperlukan karena algoritma K-Means hanya dapat memproses data numerik. Untuk memastikan skala data yang seragam, dilakukan normalisasi menggunakan operator "Normalize" dengan metode z-transformation pada atribut numerik Monthly Revenue dan Age.



GAMBAR 3.
TAHAPAN PROSES RAPID MINER

3.3 Implementasi *K-Means Clustering*

Implementasi K-Means Clustering dalam penelitian ini dilakukan menggunakan RapidMiner Studio dengan konfigurasi yang disesuaikan untuk analisis segmentasi pelanggan Netflix. Proses clustering dijalankan dengan parameter k=5 cluster dan maksimum iterasi (max runs) sebanyak 10 kali untuk memastikan konvergensi hasil yang optimal. Algoritma menggunakan measure type Euclidean Distance untuk menghitung jarak antar data point.

Hasil implementasi K-Means menghasilkan 5 cluster dengan karakteristik sebagai berikut:

a) Cluster 0: Didominasi pelanggan dengan pola Join Date dan Last Payment Date yang lebih besar dari rata-rata (>1.698%), menunjukkan pelanggan yang relatif baru bergabung.

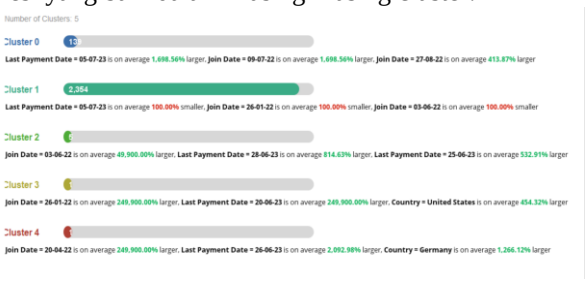
b) Cluster 1: Karakteristik utama Last Payment Date dan Join Date 100% lebih kecil dari rata-rata, mengindikasikan kelompok pelanggan lama.

c) Cluster 2: Memiliki pola Join Date yang signifikan lebih tinggi (49.900%) dan Last Payment Date yang lebih besar (814.63%), menunjukkan pelanggan aktif dengan tingkat engagement tinggi.

d) Cluster 3: Terdiri dari pelanggan dengan dominasi dari United States (454.32% lebih tinggi dari rata-rata) dan pola Join Date yang konsisten.

e) Cluster 4: Karakteristik khusus dengan pelanggan dari Germany (1.266.12% lebih tinggi) dan pola pembayaran yang teratur.

Kualitas hasil clustering divalidasi menggunakan Davies Bouldin Index dengan nilai -2.920, mengindikasikan pemisahan cluster yang baik. Nilai rata-rata within centroid distance untuk setiap cluster (-341.664, -162.674, -353.027, -29.268) menunjukkan kohesi yang baik dalam masing-masing cluster.



GAMBAR 4.
HASIL CLUSTERING K-MEANS

3.3 Evaluasi Model

Evaluasi kinerja model clustering dilakukan menggunakan Performance Vector yang menghasilkan beberapa metrik penting. Hasil evaluasi menunjukkan:

a. Average Within Centroid Distance:
Keseluruhan: -341.664
Cluster_0: -162.674
Cluster_1: -353.027
Cluster_2: -29.268
Cluster_3: -0.000
Cluster_4: -0.000

Nilai average within centroid distance yang negatif mengindikasikan jarak yang relatif dekat antara data point dengan centroid masing-masing cluster, menunjukkan kohesi yang baik dalam cluster.

b. Davies Bouldin Index:
Nilai: -2.920

Davies Bouldin Index yang bernilai -2.920 mengindikasikan kualitas clustering yang baik. Nilai negatif menunjukkan separasi antar cluster yang optimal. Semakin kecil nilai absolut Davies Bouldin Index, semakin baik kualitas cluster. Hasil ini menunjukkan bahwa model berhasil memisahkan data ke dalam cluster-cluster yang berbeda dengan baik. Berdasarkan hasil evaluasi ini, dapat disimpulkan bahwa implementasi K-Means Clustering dengan 5 cluster menghasilkan segmentasi yang baik dan dapat diandalkan untuk analisis karakteristik pelanggan Netflix.

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024

GAMBAR 5.
HASIL EVALUASI MODEL

IV. KESIMPULAN

Berdasarkan hasil penelitian segmentasi pelanggan Netflix menggunakan metode K-Means Clustering, dapat disimpulkan bahwa algoritma ini berhasil membagi 2500 pelanggan ke dalam 5 cluster dengan karakteristik yang berbeda. Kualitas clustering ditunjukkan oleh nilai Davies Bouldin Index -2.920, mengindikasikan pemisahan cluster yang optimal. Hasil clustering mengidentifikasi beberapa segmen pelanggan utama, termasuk pelanggan baru (Cluster 0), pelanggan lama (Cluster 1), pelanggan dengan engagement tinggi (Cluster 2), serta segmentasi berdasarkan lokasi geografis seperti pelanggan United States (Cluster 3) dan Germany (Cluster 4). Temuan ini memberikan pemahaman yang lebih mendalam tentang karakteristik dan perilaku pelanggan Netflix.

V. SARAN

Untuk pengembangan lebih lanjut, disarankan agar Netflix memanfaatkan hasil segmentasi ini untuk mengembangkan strategi pemasaran yang lebih personal dan terarah sesuai karakteristik masing-masing cluster. Netflix juga direkomendasikan untuk melakukan analisis berkala mengingat dinamika preferensi pelanggan yang terus berubah. Penelitian selanjutnya dapat mengeksplorasi penggunaan algoritma clustering lain seperti Hierarchical Clustering atau DBSCAN, serta menambahkan variabel analisis seperti genre tontonan favorit dan durasi menonton untuk mendapatkan insight yang lebih komprehensif. Selain itu, penerapan teknik visualisasi yang lebih advanced dapat membantu dalam interpretasi hasil clustering yang lebih baik.

DAFTAR PUSTAKA

[1] Tempo Pusat data dan analisis, “Mengenal Situs Netflix yang populer dan upaya pengaturan oleh pemerintah”, Penerbit; Tempo publishing, Jakarta. februari 2019.

[2] B. . Putri, D. . Amalia, D. Aulia, N. Salsabila, N. Hasna, and Ramadina. Syahrani, “Analisis Nilai Tambah Perusahaan Berdasarkan Business Model Canvas": Netflix dan Amazon Prime Video,” 2021.

[3] A. Wasik *et al.*, “Implementasi data mining untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1,” vol. 1, no. 2, pp. 469–479, 2024.

[4] E. C. Vidiya and G. Testiana, “Analisis Pola Pembelian di Lathansa Cafe & Ramen dengan Menggunakan Algoritma FP-Growth Berbantuan RapidMiner,” *G-Tech J. Teknol. Terap.*, vol. 7, no. 3, pp. 1118–1126, 2023, doi: 10.33379/gtech.v7i3.2739.

[5] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021, doi: 10.30813/jbase.v4i1.2729.

[6] K. McDonald and D. Smith-Rowsey, "The Netflix Effect: Technology and Entertainment in the 21st Century".*The Journal of Popular Culture.* vol. 50, no. 6, Pp 1443-1446, 2016. doi: 10.5040/9781501309410.

[7] Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al- Musa, and Edi Tohidi, “Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang,” *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 2, pp. 58–64, 2022, doi: 10.32485/kopertip.v4i2.120.

[8] D. B. Utami, “Mengenal Indonesia Melalui Netflix Original Movie,” *J. Komun.*, vol. 11, no. 1, p. 70, 2019, doi: 10.24912/jk.v11i1.4051.

[9] A. Jalil, A. Homaidi, and Z. Fatah, “Implementasi Algoritma Support Vector

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024

Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.

[10] H. Alrasyid, A. Homaidi, M. Kom, Z. Fatah, and M. Kom, "Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes," *ICORHESTECH Proceeding of Internasional Conference Of Religion, Health, Education, Science, And Technology*. vol. 1, no. 1, pp. 447–453, 2024.

Copyrigt@Jurnal Sains dan Sistem Teknologi Informasi (SANDI)
Dokumen diterima pada 18-11-2024
Dipublikasikan pada 30-11-2024